



マルコフ確率場による未観測データ補完と領域分割に関する研究

著者	片岡 駿
学位授与機関	Tohoku University
学位授与番号	11301甲第15930号
URL	http://hdl.handle.net/10097/58712

論文内容の要約

第1章 序論

近年の計算機の高速化と大容量通信環境の普及に伴い、以前の時代では想像もできないような高性能なIT 機器が我々の生活に存在するようになった。スマートフォンの普及や家電製品のIT 化がその一例である。こうした新しいIT 機器やサービスの出現を受けて、データや情報というのが急速にその価値を高めている。我々はこれらの機器を使用することで購買履歴や通話ログ、GPS情報などのデータを無意識のうちに大量に作り出しているのである。このようなデータはビッグデータと呼ばれ、多種多様な情報を陰に含んでいることが多く、そこから我々の想定した情報のみならず想定外の情報をも抽出できる可能性がある。そのため、これからのビッグデータ時代ではこのような大量のデータを最大限に利用し、いかに重要な情報を引き出すかが、今後の経営戦略を考える上での重要なポイントであると考えられる。確率的情報処理はこのようなビッグデータの活用において注目されている技術の一つである。大量のデータから必要なデータを抽出するにはデータの持つノイズなどの不確実さに起因するランダムネスやデータの各要素間に存在する複雑な相関の取扱いを考える必要がある。確率的情報処理とは情報処理を行うシステムを確率モデルとして表現し、データの不確実さや要素間の相関を確率論の数理的な枠組みの下で取り扱うことを考える情報処理の方法である。

確率的な枠組みの下で情報処理を行っていくにあたって、重要となるのは情報処理を行うシステムを表現する確率モデルの設計と実際に情報処理を行うアルゴリズムの開発、そして作成された情報処理システムが実際にどの程度の情報処理能力を持っているのかを見積もる統計的性能評価である。しかしながら、実際に確率的情報処理を行うことを考えると、その大きな計算量が重大な問題となる。確率モデルの設計やアルゴリズムの開発、統計的性能評価は統計学や統計的機械学習理論、情報統計力学の枠組みの中で具体的な定式化が既になされているが、これらの計算は多重和などの計算を必要とし現実的な時間内では計算不可能である。そのため、確率的な枠組みの下で情報処理を行うためには確率モデルの設計およびアルゴリズムの開発、統計的性能評価の計算を現実時間内で可能とする近似計算法を開発する必要がある。

本論文の目的は確率分布の周辺確率や統計量を固定点方程式の計算により近似的に計算する統計力学的近似計算法を利用することでこの計算時間の問題を回避する具体的な計算法を提案することである。具体的にはガウシアングラフィカルと呼ばれる比較的扱いやすい確率モデルを使用し、画像補修問題と交通量の補完問題を題材として確率モデルの設計やアルゴリズムの開発、統計的性能評価の問題が統計力学的近似計算法を利用することで現実的な時間内で近似的に計算可能なことを示す。さらに、完全結合型の条件付き確率場に対しても統計力学的近似計算法による具体的な推論アルゴリズムを提案することで、ガウシアングラフィカルモデルにおいて提案した計算法をより複雑で最新の確率モデルに対して拡張することを目指す。

第2章 確率的情報処理

大規模なデータを扱う確率的な情報処理システムの実現において重要となる確率モデルの設計法と確率モデルからの推定、統計的性能評価について概説する。情報処理の諸問題はすべて手元にある情報から知りたい情報を抽出または推論する問題と解釈することができる。確率的情報処理の観点では、このことは事後確率分布からの確率的推論問題として定式化することができる。

本章では事後確率分布の設計指針が識別モデルと生成モデルの二通りの方法があることを説明し、事後確率分布からの情報の抽出法として最大事後確率と最大周辺事後確率という2つの推論基準を与えた。さらに、情報処理システムの統計的性能評価が生成モデルの方法のように観測過程と事前確率分布を設計し、設計された確率分布に関する期待値計算で与えられることを説明した。

第3章 統計力学的近似計算法

確率的情報処理で行われる確率モデルの構築やアルゴリズムの開発、統計的性能評価は多くの場合計算困難の問題を抱えており、現実的な時間内では計算不可能である。そのため、現実的な時間内で計算するためには何らかの近似計算を導入する必要がある。

本章では本論文で使用する固定点方程式の方法である統計力学的近似計算法について説明する。この方法は試行分布をその引数とする変分自由エネルギーの最小化に基づく方法であり、実際の確率分布と試行分布の間のKL情報量を最小にする試行分布で実際の確率分布を置き換える方法と解釈することができる。この変分自由エネルギーの最小化には試行分布の構造を固定する方法と試行分布の統計量を固定する方法の二種類があり、平均場近似やTAP近似と呼ばれる近似計算法がこれらの方法から導出されることを述べた。また、線形応答の関係式に注目することで、平均場近似から感受率と呼ばれる統計量を計算する近似計算法である感受率伝搬が導出されることを説明した。

第4章 確率モデルに基づく画像補修アルゴリズムとその性能評価

画像補修問題とは画像から物体を除去したときに残る背後の穴等の画素値の欠損を埋める画像処理の手法であり、ガウシアングラフィカルモデルに基づく効果的な画像補修アルゴリズムが既に提案されている。本章では第3章で説明した統計力学的近似計算法を利用することで画像補修問題の確率モデルの統計的性能評価が近似的に計算可能なことを示す。

ガウシアングラフィカルモデルに基づく画像補修アルゴリズムは画像のグラフ構造を正方格子と仮定し、欠損している画素の推定値を隣接する欠損していない画素の値や欠損している画素の推定値の値が滑らかにつながるような画像を欠損画像からの推定画像とする方法であり、画像の具体的な推定は平均場近似による固定点方程式を計算することで行われる。

本研究では、この確率モデルの統計的性能をランダム欠損に対する元の画像と推定画像の間の平均二乗誤差の期待値として定式化を行った。平均場近似による欠損画素の推定値の分布を平均場近似の固定点方程式を利用した積分方程式の解として表し、画像補修問題の確率モデルの統計的性能評価を計算する近似的計算法を構築した。実際に標準画像に対して確率モデルの統計的性能評価を行い、ガウシアングラフィカルモデルに基づく画像補修法のランダム欠損に対する統計的性能を明らかにした。

第5章 統計力学的近似計算法に基づくEMアルゴリズムの改良

第4章で解説したガウシアングラフィカルモデルに基づく画像補修アルゴリズムは欠損した画素値の推定を平均場近似による固定点方程式の計算問題に置き換えることで欠損した画素値の推定を行うアルゴリズムであった。しかし、このアルゴリズムはパラメータを含む確率モデルから設計された計算法であるため、アルゴリズムを適用する前にこのパラメータを適切に与える必要がある。

本章ではEMアルゴリズムと感受率伝搬法を組み合わせることで、画像補修問題の確率モデルのパラメータを決定するパラメータ推定法を提案する。EMアルゴリズムとは確率モデルに隠れ変数を含む時に効果的に確率モデルのパラメータを推定する方法であり、画像補修問題の場合には欠損している画素の画素値が確率モデルの隠れ変数に対応する。この方法では、パラメータの推定はQ関数と呼ばれるパラメータに関する関数を繰り返し最大化することで達成され、最大化に必要なQ関数の勾配は確率モデルの統計量を使って計算される。

本研究ではQ関数の勾配の計算に平均場近似と感受率伝搬法を使用し、固定点方程式の計算を途中で打ち切ることで、EMアルゴリズムのQ関数の繰り返し最大化を効果的に計算するアルゴリズム

を提案した．実際に標準画像を用いた数値実験により，提案法により画像補修アルゴリズムが良好な結果を与えるパラメータが推定できることを明らかにした．

第6章 確率モデルに基づく交通情報の補完法

本章では第4章で説明したガウシアングラフィカルモデルに基づく画像補修法を拡張して，交通密度の観測されていない道路の交通密度を補完する方法を提案する．渋滞予測や旅行時間といった交通予測の高精度な情報をドライバーに提供するためには，事前に交通密度や平均速度などの交通情報を広範囲かつ高精度に収集する必要がある．しかしながら，交通情報を収集する交通センサーをすべての道路に設置することは事実上不可能であり，実際に広範囲の交通情報を収集することはできない．そのため，交通予測を行うためにはセンサーの設置されていない道路の交通情報を補完する必要がある．

ガウシアングラフィカルモデルに基づく画像補修法は欠損していない画素の画素値と欠損している画素の画素値が滑らかにつながるように画像を推定するものであった．この仮定は道路の交通密度や平均速度に対しても十分適用できるものと思われる．

本研究では道路交通網の各道路を頂点とし，交差点でつながっている道路に対応する頂点同士を辺でつなぐことで，交通密度を表す確率モデルのグラフ構造を定義した．画像補修問題の確率モデルと同様に隣接間の交通密度が滑らかにつながる確率が高くなるような相互作用項を採用し，幹線道路等の主要な道路は普段から交通密度が大きい等といった道路固有の特徴を線形なバイアス項で表すことで交通密度を表すパラメトリックな確率モデルの構築を行った．この確率モデルのパラメータは，過去に集められた交通情報の利用により疑似的な欠損の無い交通密度データが作成可能であると仮定することで，統計的機械学習理論の正則化学習の方法と平均場近似を組み合わせることで決定される．実際の交通密度の補完は画像補修問題と同様に平均場近似の固定点方程式を計算することで行われる．仙台市を模した大規模なシミュレーションデータを使用して実際に数値実験を行い，提案した交通密度の補完法が単純な逐次的補完法と比べて良好な補完結果を与えることを明らかにした．

第7章 完全結合型条件付き確率場に基づく画像の領域分割アルゴリズム

第4章から第6章までで得られた結果によってガウシアングラフィカルモデルに基づいた未観測データの補完法に対するパラメータの決定による確率モデルの設計やアルゴリズムの開発，統計的性能評価の具体的な計算法が統計力学的近似計算法を利用することで構成できることが明らかとなった．本章では完全結合型条件付き確率場を使用して画像の領域分割アルゴリズムを提案することで，より複雑で最新の確率モデルに対しても確率的情報処理の諸計算に対する具体的な計算法を構築することを目指す．

画像の領域分割とは，ある基準に従って画像中から一緒になる画素のグループを見つけることである．領域分割を行うことで画像を特徴のあるグループに分割することができるため，パターン認識等のコンピュータビジョンの様々な分野で利用されており，画像処理の研究において古くから盛んに研究されている問題の一つである．

近年，物体認識の研究で完全結合型の確率モデルが物体全体の情報を上手く捉えるのに有効であるとの報告がなされている．本研究では確率モデルのグラフ構造を完全グラフとし，画像中の画素間の距離が大きくなるごとにその相互作用の強さが指数的に減少するような相互作用項と混合ガウス分布を利用して画素値の近さで領域の優先度を変化させるバイアス項を採用することで，完全結合型の条件付き確率場を用いた領域分割の確率モデルを構築した．具体的な領域分割の計算法はTAP近似の固定点方程式を計算することで達成される．領域分割用のデータセットを用いて実際に領域分割を行い，提案した領域分割法により混合ガウス分布のソフトクラスタリングの結

果よりも良好な分割結果が得られることを明らかにした.

第8章 結論

本論文では未観測データの補完問題と領域分割問題の両方に対して対する新しい確率的計算モデルの提案と統計力学的近似計算法を利用した確率的情報処理の諸問題に対する具体的な計算法の構築を行った. 確率モデルを利用して大規模な情報処理を行うためには統計量の計算等の計算困難として知られる計算を何度も行う必要がある. そのため, この計算困難の問題を回避して適切に情報処理を行う具体的な近似計算法の構築が実際の情報処理の現場では求められている.

このような状況に着目し, 本論文ではガウシアングラフィカルモデルという比較的扱いやすい確率モデルを用いた未観測データの補完問題を扱い, パラメータの推定という確率モデルの設計にかかわる問題や実際に推論を行うアルゴリズム, さらに確率モデルの統計的性能評価といった確率的情報処理の諸問題に対する具体的な近似計算法がすべて固定点方程式を計算する統計力学的近似計算法を利用することで構成可能であることを明らかにし, 数値実験によって良好な結果が得られることを確認した. また, 完全結合型の条件付き確率場というより複雑で最新の確率モデルに対しても同様な近似計算法を構成することを目指し, 完全結合型条件付き確率場を利用した領域分割の確率モデルを構成した. この確率モデルに対して統計力学的近似計算法を利用した推論アルゴリズムを提案し, 数値実験により良好な結果が得られることを明らかにした.